

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2002, 00:09:25 ; Search time 8498.8 Seconds
(Without alignments)
29.081 Million cell updates/sec

Title: US-09-851-670-15

Perfect score: 23
Sequence: 1 aacgtgctgcgtccacagaca 23

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 111874

Minimum DB seq length: 0
Maximum DB seq length: 60

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estfun: *
2: em_esthum: *
3: em_estin: *
4: em_estom: *
5: em_estpl: *
6: em_estba: *
7: em_estro: *
8: em_estov: *
9: em_htc: *
10: qd_estl: *
11: qd_est2: *
12: qd_htc: *
13: qd_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pln: *
18: em_gss_pro: *
19: em_gss_rtd: *
20: em_gss_vrt: *
21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 13.6 | 59.1 | 24 | 13 | AZ490697 IM0323016 |
| 2 | 13.4 | 58.3 | 46 | 10 | A1416932 |
| 3 | 13.2 | 57.4 | 48 | 10 | A1416932 |
| 4 | 13 | 56.5 | 50 | 13 | AZ341480 IM0073D24 |
| 5 | 12.8 | 55.7 | 50 | 10 | AZ857216 2M0161M13 |
| 6 | 12.8 | 55.7 | 50 | 10 | AU105784 AU105784 |
| 7 | 12.8 | 55.7 | 50 | 10 | AU105785 AU105785 |
| 8 | 12.6 | 54.8 | 34 | 13 | AZ767937 |
| 9 | 12.6 | 54.8 | 41 | 13 | AZ412970 |
| 10 | 12.6 | 54.8 | 44 | 13 | TA121E040 |
| 11 | 12.6 | 54.8 | 45 | 13 | AQ026252 |
| 12 | 12.6 | 54.8 | 58 | 10 | A1290333 |

| | | | | | | |
|----|------|------|----|----|------------|--------------------|
| 13 | 12.6 | 54.8 | 59 | 11 | R71912 | R71912 yf84c05.s1 |
| 14 | 12.4 | 53.9 | 50 | 10 | AU104389 | AU104389 AU104389 |
| 15 | 12.2 | 53.0 | 26 | 13 | AZ514624 | AZ514624 IM0361P19 |
| 16 | 12.2 | 53.0 | 50 | 10 | AU105302 | AU105302 AU105302 |
| 17 | 12.2 | 53.0 | 52 | 13 | AZ440200 | AZ440200 IM0231I07 |
| 18 | 12.2 | 53.0 | 53 | 10 | AW692002 | AW692002 NF051D01S |
| 19 | 12 | 52.2 | 25 | 13 | TA934FE01P | TA934FE01P |
| 20 | 12 | 52.2 | 41 | 13 | TA335H09P | TA335H09P |
| 21 | 12 | 52.2 | 45 | 13 | AZ307664 | AZ307664 IM0009J21 |
| 22 | 12 | 52.2 | 50 | 10 | AU106853 | AU106853 AU106853 |
| 23 | 12 | 52.2 | 50 | 10 | AU107096 | AU107096 AU107096 |
| 24 | 12 | 52.2 | 58 | 10 | AA917315 | AA917315 0455a02.s |
| 25 | 11.8 | 51.3 | 33 | 13 | TA215H02P | TA215H02P |
| 26 | 11.8 | 51.3 | 49 | 11 | BF017790 | BF017790 |
| 27 | 11.8 | 51.3 | 50 | 13 | AZ949090 | AZ949090 |
| 28 | 11.8 | 51.3 | 51 | 11 | BF224930 | BF224930 |
| 29 | 11.8 | 51.3 | 51 | 13 | AZ768129 | AZ768129 IM0505FP3 |
| 30 | 11.8 | 51.3 | 57 | 13 | AZ803892 | AZ803892 2M0064F05 |
| 31 | 11.8 | 51.3 | 58 | 11 | N44616 | N44616 YJ30a12.r1 |
| 32 | 11.8 | 51.3 | 58 | 11 | T48124 | T48124 Yb25b05.r1 |
| 33 | 11.8 | 51.3 | 59 | 10 | BE022784 | BE022784 sm88a07.y |
| 34 | 11.6 | 50.4 | 37 | 10 | A1873935 | A1873935 vm44d08.x |
| 35 | 11.6 | 50.4 | 37 | 13 | AZ430297 | AZ430297 IM0214A1S |
| 36 | 11.6 | 50.4 | 37 | 13 | AZ806455 | AZ806455 2M0068E04 |
| 37 | 11.6 | 50.4 | 45 | 13 | AZ458735 | AZ458735 1M0263P04 |
| 38 | 11.6 | 50.4 | 48 | 13 | AZ828038 | AZ828038 2M0104L17 |
| 39 | 11.6 | 50.4 | 50 | 10 | AU104153 | AU104153 AU104153 |
| 40 | 11.6 | 50.4 | 50 | 10 | AU104155 | AU104155 AU104155 |
| 41 | 11.6 | 50.4 | 50 | 10 | AU104740 | AU104740 AU104740 |
| 42 | 11.6 | 50.4 | 52 | 13 | AZ311262 | AZ311262 IM0026B12 |
| 43 | 11.6 | 50.4 | 55 | 10 | BE374434 | BE374434 601227409 |
| 44 | 11.6 | 50.4 | 60 | 13 | AZ694939 | AZ694939 TE-389-61 |
| 45 | 11.4 | 49.6 | 32 | 13 | AZ478463 | AZ478463 IM0298J05 |

ALIGNMENTS

| | | | | | |
|------------|---|--------------------|-------|---------|----------------------|
| RESULT 1 | AZ490697/c | 24 bp | DNA | GSS | 05-OCT-2000 |
| LOCUS | IM0323016R | Mouse 10kb plasmid | UUC1M | library | Mus musculus genomic |
| DEFINITION | clone UUC1M0323016 R, DNA sequence. | | | | |
| ACCESSION | AZ490697 | | | | |
| VERSION | AZ490697.1 | GI:10661682 | | | |
| KEYWORDS | GSS. | | | | |
| SOURCE | house mouse. | | | | |
| ORGANISM | Mus musculus | | | | |
| REFERENCE | Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R. | | | | |
| AUTHORS | Mouse whole genome scaffolding with paired end reads from 10kb | | | | |
| TITLE | Plasmid inserts | | | | |
| JOURNAL | Unpublished (2000) | | | | |
| COMMENT | Contact: Robert B. Weiss University of Utah Genome Center Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert length: 10000 Std Error: 0.00 Plate: 0323 row: 0 column: 16 Seq primer: CACACAGAAACACCTTGACC Class: plasmid ends High quality sequence stop: 24. Location/Qualifiers 1. 24 | | | | |
| FEATURES | | | | | |
| SOURCE | | | | | |

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U081M0323016"
 /clone_lib="Mouse 10kb plasmid U081M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/shares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT
 ORIGIN
 8 a 7 c 3 g 6 t

Query Match 59.1%; Score 13.6; DB 13; Length 24;
 Best Local Similarity 80.0%; Pred. No. 1.3e+04;
 Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 4 gtgtcggctccctcagagaca 23
 ||||| ||| ||||| |||
 Db 23 GGTGCTGCTGATCATCATCA 4

RESULT 2
 A1416932/c 46 bp mRNA EST 01-DEC-1999
 LOCUS A1416932
 DEFINITION sal1f09.Y1 Gm-cl004 GLYCINE max CDNA clone GENOME SYSTEMS CLONE ID: Gm-cl004-354.5' similar to SM-TTIPM_PBA P25794 WATER-STRESS INDUCED TOMOPLAST INTRINSIC PROTEIN;; mRNA sequence.
 ACCESSION A1416932.1 GI:4260436
 VERSION A1416932
 KEYWORDS EST.
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
 1 (bases 1 to 46)

REFERENCE
 AUTHORS Shoemaker, R., Keim, P., Vothkin, L., Erpelding, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harrey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
 TITLE Public Soybean EST Project
 JOURNAL Unpublished (1999)
 COMMENT Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 for further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)

427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 Trace considered overall poor quality
 Possible reversed clone: similarity on wrong strand
 Seq primer: -40RP from Gibco
 High quality sequence stop: 1.
 Location/Qualifiers
 1. 46
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl004-354"
 /clone_lib="Gm-cl004"
 /tissue_type="root"
 /lab_host="XL10-Gold"
 /note="Vector: pBluescript II XRI; Site_1: EcoRI; Site_2: XhoI; Root CDNA. The mRNA was isolated from entire roots of 8 day old 'Williams' seedlings which were propagated on paper towels with distilled water. Stratagene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA is hemimethylated. Stratagene's first-strand synthesis primer was used (GAGGAGAGAGAGAGAGACTGCTCGAG(T)-18). After second-strand synthesis, the cDNA ends were 'polished' with clone Pfu DNA polymerase, ligated to EcoRI adaptors, and phosphorylated. The XhoI site within the first-strand synthesis primer was restricted by digestion with XhoI; all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500bp cutoff, using GIBCOBRL Life Technologies' cDNA Size Fractionation column. The column eluent was then ligated into Stratagene's pBluescript II XRI Predigested vector (pBluescript II SK(+)) that had been digested with EcoRI and XhoI, and phosphorylated). Both the white and blue colonies appear to contain recombinant plasmids with cDNA inserts. Blue colonies 9n-15 have been sequenced, and possess putative cDNA inserts. This library was constructed by Dr. Paul Keim & Virginia H. Coryell, Department of Biology, Box5640, Northern Arizona University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr. Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax: 520-523-7500, email: paul.keim@na.u.arizona.edu, virginia.coryell@na.u.arizona.edu"

BASE COUNT
 ORIGIN
 11 a 7 c 20 g 8 t

Query Match 58.3%; Score 13.4; DB 10; Length 46;
 Best Local Similarity 73.9%; Pred. No. 1.8e+04;
 Matches 17; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 1 aacgtgtcggctccctcagagaca 23
 ||| ||| ||| ||| ||| |||
 Db 43 AACCTGTCGCTCCACACAGACA 21

RESULT 3
 A2341480/c 48 bp DNA GSS 29-SEP-2000
 LOCUS A2341480
 DEFINITION IM0073D2R Mouse 10kb plasmid U081M library Mus musculus genomic clone U081M0073D2 R, DNA sequence.
 ACCESSION A2341480
 VERSION A2341480.1 GI:10417773
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 48)

REFERENCE
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausern, A. and Wright, D., Weiss, R.

TITLE
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL
Unpublished (2000)
COMMENT
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0073 row: D column: 24
Seq primer: CACACAGCAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 48.
Location/Qualifiers
1. .48
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C1M0073D24"
/clone_lib="Mouse 10kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (911473211419b1AF129072.1), a copy number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT
11 a 16 c 13 g 8 t
ORIGIN

Query Match 57.4%; Score 13.2; DB 13; Length 48;
Best Local Similarity 83.3%; Pred. No. 2.3e+04;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 cgtgtcggtcctcagag 20
1 ||| |||||
Db 45 CTTGTATGCTCCTCAGAG 28

RESULT 4
A2857216 50 bp DNA GSS 21-FEB-2001
LOCUS A2857216
DEFINITION clone U06C2M0161M13 R, DNA sequence.
ACCESSION A2857216
VERSION A2857216.1 GI:13048831
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 50)
Dunn, D., Aoyagi, A., Barber, M., Beacom, T., Duval, B., Hamill, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.

TITLE
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL
Unpublished (2000)
COMMENT
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0161 row: M column: 13
Seq primer: CACACAGCAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 50.
Location/Qualifiers
1. .50
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C2M0161M13"
/clone_lib="Mouse 10kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (911473211419b1AF129072.1), a copy number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT
8 a 17 c 12 g 13 t
ORIGIN

Query Match 56.5%; Score 13; DB 13; Length 50;
Best Local Similarity 76.2%; Pred. No. 2.9e+04;
Matches 16; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 cgtgtcggtcctcagag 23
1 ||||| |||||
Db 17 CCTGTCTGTCTCCCTGGACA 37

RESULT 5
AUI05784 50 bp mRNA EST 05-APR-2001
LOCUS AUI05784
DEFINITION Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
HRC03176, mRNA sequence.
ACCESSION AUI05784
VERSION AUI05784.1 GI:1355305
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 50)
Suzuki, Y., Tsunoda, T., Taira, H., Mizushima-Sugano, J., Sese, J., Hata,
H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., Morishita, S., Okubo

TITLE K., Suyama, A. and Sugano, S.
Fine Structural Analysis of transcription start sites of human
mRNAs using full-length enriched and 5'-end enriched cDNA libraries
Unpublished (2001)
JOURNAL
COMMENT Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: yusuzuki@ims.u-tokyo.ac.jp
Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano
, S. Construction and characterization of a full length-enriched and
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES
source
1. 50
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HRC03176"
/clone.lib="Sugano Homo sapiens cDNA library"

BASE COUNT
ORIGIN
7 a 13 c 25 g 5 t

Query Match
Best Local Similarity 55.7%; Score 12.8; DB 10; Length 50;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 8 gcggtccagagaca 23
|||||
Db 41 GCGGTCTCCGAGCCA 26

RESULT 6
LOCUS AUI05785 50 bp mRNA EST 05-APR-2001
DEFINITION AUI05785 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
HS106841, mRNA sequence.
ACCESSION AUI05785
VERSION AUI05785.1 GI:1355306
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 50)
Suzuki, Y., Tsunoda, T., Taira, H., Mizushima-Sugano, J., Sese, J., Hata
, H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., Morishita, S., Okubo
, K., Suyama, A. and Sugano, S.
Fine Structural analysis of transcription start sites of human
mRNAs using full-length enriched and 5'-end enriched cDNA libraries
Unpublished (2001)
JOURNAL
COMMENT Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: yusuzuki@ims.u-tokyo.ac.jp
Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano
, S. Construction and characterization of a full length-enriched and
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES
source
1. 50
Location/Qualifiers

BASE COUNT
ORIGIN
6 a 13 c 25 g 6 t

Query Match
Best Local Similarity 55.7%; Score 12.8; DB 10; Length 50;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 8 gcggtccagagaca 23

Db 45 GCGGTCTCCGAGCCA 30
|||||
RESULT 7
LOCUS AUI07337 50 bp mRNA EST 05-APR-2001
DEFINITION AUI07337 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
CAS00717, mRNA sequence.
ACCESSION AUI07337
VERSION AUI07337.1 GI:13556858
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 50)
Suzuki, Y., Tsunoda, T., Taira, H., Mizushima-Sugano, J., Sese, J., Hata
, H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., Morishita, S., Okubo
, K., Suyama, A. and Sugano, S.
Fine Structural analysis of transcription start sites of human
mRNAs using full-length enriched and 5'-end enriched cDNA libraries
Unpublished (2001)
JOURNAL
COMMENT Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: yusuzuki@ims.u-tokyo.ac.jp
Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano
, S. Construction and characterization of a full length-enriched and
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES
source
1. 50
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CAS00717"
/clone.lib="Sugano Homo sapiens cDNA library"

BASE COUNT
ORIGIN
3 a 12 c 22 g 13 t

Query Match
Best Local Similarity 55.7%; Score 12.8; DB 10; Length 50;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 cgtgtcggtccag 18
|||||
Db 25 GCGGCGGTCTCAG 40

RESULT 8
LOCUS A2767937 34 bp DNA GSS 16-FEB-2001
DEFINITION IM056708R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M056708 R, DNA sequence.
ACCESSION A2767937
VERSION A2767937.1 GI:12886546
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 34)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamll, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly
, M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausen, A.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
JOURNAL
COMMENT Contact: Robert B. Weiss
University of Utah
University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA

Tel: 801 585 5606
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0567 row: F column: 08

Seq primer: CACACAGAAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 34.

FEATURES

Location/Qualifiers

1..34

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC1M0567F08"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pMD42 (g11473211419b1AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

4 a 9 c 10 g 11 t

BASE COUNT
ORIGIN

Query Match 54.8%; Score 12.6; DB 13; Length 34;

Best Local Similarity 78.9%; Pred. No. 4.1e+04;

Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 5 tgtgcgctctcagagaca 23

DB 25 TGTGAGTCTCTGACACACA 7

RESULT 9
AZ412970/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center

University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA

Tel: 801 585 5606
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0186 row: G column: 15

Seq primer: CACACAGAAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 41.

FEATURES

Location/Qualifiers

1..41

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC1M0186G15"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pMD42 (g11473211419b1AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

14 a 10 c 8 g 9 t

BASE COUNT
ORIGIN

Query Match 54.8%; Score 12.6; DB 13; Length 41;

Best Local Similarity 78.9%; Pred. No. 4.2e+04;

Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 4 gtgtgcgctctcagagac 22

DB 20 GTATGTGCTCTGACAC 2

RESULT 10
TA121E040

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

Trypanosoma brucei genome sequencing
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus,
Hinxton, Cambridgeshire CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk

COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of *Trypanosoma brucei* (TREU927/4 Gurat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available at <http://www.sanger.ac.uk/Projects/T-brucei/>.

FEATURES
Source
1..44
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="121e04"

BASE COUNT 10 a 15 c 8 g 11 t

ORIGIN

Query Match 54.8%; Score 12.6; DB 13; Length 44;
Best Local Similarity 78.9%; Pred. No. 4.3e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 aacgttcgcgtcctcaga 19
||||| 11 11111 1
Db 18 AACGTGCGCAGCCTCAAA 36

RESULT 11
A0026252 45 bp DNA GSS 30-JUN-1998
LOCUS l(3)l330 Drosophila melanogaster P lethal line Drosophila
DEFINITION melanogaster genomic Sequence recovered from 5' end of P element,
DNA sequence.
A0026252
A0026252.1 GI:3266537
GSS.
fruit fly.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 45)
Spradling, A.C., Stern, D., Beaton, A., Rehm, E.J., Laverly, T., Mozer
, N., Mista, S. and Rubin, G.M.
The BOP gene disruption project: Single P element insertions
mutating 30% of Drosophila autosomal genes
Unpublished (1998)
Contact: Gerald Rubin
Berkeley Drosophila Genome Project
University of California, Berkeley
LSA Building, Berkeley, CA 94720-3200, USA
Fax: 5106439947
Email: gerry@fruitfly.berkeley.edu
Sequence recovery method was inverse PCR.
Sequence orientation is forward strand relative to 5' end of P
element

The P element insertion position is base 038 in the 45 bases. This
insertion position refers to the first base of the 8 base target
recognition sequence.
Class: transposon-tagged.
Location/Qualifiers
1..45
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone.lib="Drosophila melanogaster P lethal line"
/note="Inverse PCR was performed on Drosophila
melanogaster strains each of which contains a single P

FEATURES
Source
1..45
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone.lib="Drosophila melanogaster P lethal line"
/note="Inverse PCR was performed on Drosophila
melanogaster strains each of which contains a single P

transposable element insertion that is thought to cause
either lethality or sterility. The resultant fragment for
each strain was directly sequenced to determine the
genomic sequence at the site of insertion. Details of the
protocols used can be found at
http://fruitfly.berkeley.edu/P-disrupt/inverse_pcr.html.

BASE COUNT 9 a 13 c 16 g 7 t

ORIGIN

Query Match 54.8%; Score 12.6; DB 13; Length 45;
Best Local Similarity 78.9%; Pred. No. 4.3e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 4 gtgtcggctcctcagagac 22
||||| 11 1111
Db 13 GCGTGGCGTGGCAAGAC 31

RESULT 12
A1290333 58 bp mRNA EST 29-JAN-1999
LOCUS qm02d02.x1 Soares_NhMPU.S1 Homo sapiens cDNA clone IMAGE:1880643
DEFINITION 3' similar to SW:THCC_HUMAN 000154 CYTOSOLIC ACT1 COENZYME A
THIOESTER HYDROLASE ; mRNA sequence.
A1290333
A1290333.1 GI:3933107
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 58)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Trace considered overall poor quality
Insert length: 1027 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 1.
Location/Qualifiers
1..58
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone.lib="IMAGE:1880643"
/clone.lib="Soares_NhMPU.S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2bhm, pregnant uterus
NbHPV, and fetal heart NbH19W) were mixed, and ss circles
were used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260732-265223,
340488-345479, and 484488-489479."

BASE COUNT 9 a 17 c 16 g 16 t

ORIGIN

Query Match 54.8%; Score 12.6; DB 10; Length 58;
Best Local Similarity 78.9%; Pred. No. 4.6e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 aacgtgagcgtctcagag 20
|||||
Db 14 ACCTGAGCCTTCACAG 32

RESULT 13
R71912 59 bp mRNA EST 02-JUN-1995
LOCUS R71912
DEFINITION yj64c05.s1 Soares breast 2ndHbst Homo sapiens cDNA clone
IMAGE:155432 3' similar to gb:X69150.40S RIBOSOMAL PROTEIN S18
(HUMAN); mRNA sequence.

ACCESSION R71912
VERSION R71912.1 GI:845944
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 59)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rikkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
The Washu-Merck EST Project
Unpublished (1995)
CONTACT: Wilson R
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 2249
High quality sequence starts: 1
High quality sequence stops: 1
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Insert Length: 2249 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 1.
Location/Qualifiers

FEATURES
source 1..59
/organism="Homo sapiens"
/db_xref="GDB:573273"
/db_xref="taxon:9606"
/clone="IMAGE:155432"
/clone_lib="Soares breast 2ndHbst"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: breast; Vector: pTV73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTCGAGCGCCGCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of a modified pTV73 vector (Pharmacia).
Library went through one round of normalization to a Cot =
230. Library constructed by Bento Soares and M. Fatima
Bonaldo."

BASE COUNT 9 a 18 c 17 g 10 t 5 others
ORIGIN

Query Match 54.8%; Score 12.6; DB 11; Length 59;
Best Local Similarity 78.9%; Pred. No. 4.6e+04;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 gtgtgagcgtctcagag 22
|||||
Db 6 GTGTGCTGCGCTCGACAC 24

RESULT 14
AUI04389 50 bp mRNA EST 05-APR-2001
LOCUS AUI04389/c
DEFINITION AUI04389 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
HEP22267; mRNA sequence.

ACCESSION AUI04389
VERSION AUI04389.1 GI:13553910
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 50)
Suzuki, Y., Tsunoda, T., Taira, H., Mizushima-Sugano, J., Sese, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., Morishita, S., Okubo, K., Suyama, A. and Sugano, S.
Fine structural analysis of transcription start sites of human
mRNAs using full-length enriched and 5'-end enriched cDNA libraries
Unpublished (2001)
CONTACT: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: yusuzuki@ims.u-tokyo.ac.jp
Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano, S.
Construction and characterization of a full length-enriched and
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
Location/Qualifiers

FEATURES
source 1..50
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEP22267"
/clone_lib="Sugano Homo sapiens cDNA library"

BASE COUNT 17 a 8 c 10 g 15 t
ORIGIN

Query Match 53.9%; Score 12.4; DB 10; Length 50;
Best Local Similarity 72.7%; Pred. No. 5.5e+04;
Matches 16; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 aacgtgagcgtctcagag 22
|||||
Db 31 AACCTTGGCGTCTTCAGTTAC 10

RESULT 15
A2514624 26 bp DNA GSS 05-OCT-2000
LOCUS A2514624/c
DEFINITION 1M0361P19F Mouse 10kb plasmid U0GC1M library Mus musculus genomic
clone U0GC1M0361P19 F, DNA sequence.

ACCESSION A2514624
VERSION A2514624.1 GI:10695940
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 26)
Dunn, D., Aoyagi, A., Barber, M., Becorn, T., Duval, B., Hamll, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausen, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
CONTACT: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606

Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0361 row: P column: 19
 Seq primer: CGTGTAAACGACGCCACAGT
 Class: plasmid ends
 High quality sequence stop: 26.

FEATURES

source

1. 26
 Location/Qualifiers

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUC1M0361P19"
 /clone_lib="Mouse 10kb plasmid UUC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g1147321141gblAF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT
 ORIGIN

3 a 7 c 13 g 3 t

Query Match

53.0%; Score 12.2; DB 13; Length 26;

Best Local Similarity 82.4%; Pred. No. 5.9e+04;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 cgtgtgcggtcctcaga 19

||||||| 11 11 1

DB 25 CGTGTGCGGCCCCACACA 9

Search completed: March 9, 2002, 00:09:27
 Job time: 11043 sec